PREDICTION MARKETS: PRACTICAL EXPERIMENTS IN SMALL MARKETS AND BEHAVIOURS OBSERVED

Jed D. Christiansen

This paper discusses a series of prediction markets created and operated in the summer of 2006 to measure calibration and behaviour of small-scale prediction markets. The research finds that small markets are very well calibrated and determines a potential minimum threshold of participation to ensure well-calibrated results. The results also established the markets as very efficient at predicting small probabilities.

Behavioural aspects of markets are also examined. Trader behavioural types are assessed and categorised; while a small group of traders were extremely active, over half of all traders rarely traded. Market manipulation is examined and found to be occasionally effective, though only in very small markets. Finally, incentives to trade are discussed; these markets were effective with no incentives for trading at all.

1. INTRODUCTION

Few "real-world" experiments have been conducted to test the efficiency and calibration of prediction markets; the majority of articles cite predictions from the Hollywood Stock Exchange, the Iowa Electronic Markets, and TradeSports' sports and events markets.[1] All of these markets have high volumes of trades and aim to predict popular current events. The national or international scope means that the general public participating in these prediction markets usually does not face any knowledge barriers; on average, a person knows enough about these topics to trade in predictions of elections or the success of popular films.

The aim of this research was to create prediction markets in a field with no established "experts" and with little popular attraction. Such a market would resemble potential business problems, where there is a significant knowledge barrier and unexperienced traders are the norm. Rowing was chosen to be the topic for these prediction markets as it is a popular sport in the United Kingdom, but one that has essentially no press of events outside of the national team news. It is also a sport where knowledge is generally known only to participants and localised by geography, and one where there is no method or forum for aggregation of opinions.

First, calibration of the markets is examined in whole. Subsets of that data are then further examined to asses how the number of traders involved in a market affects its calibration. Behaviours exhibited by traders are also examined and categorised, as are incentives to participation since a market needs sufficient participation to succeed. As a market needs sufficient participation to success, the issues around providing incentives to traders are examined. Finally, these results are synthesised into recommendations for future use of prediction markets in the business sector. This paper demonstrates that small-scale prediction markets, where knowledge barriers

and privacy considerations dramatically limit the pool of participants, can effectively generate reliable information from which any organisation can make better decisions.

2. PREDICTION MARKETS IN APPLICATION

In 1988, the University of Iowa developed simple markets to trade unique assets: contracts that paid out according to the winners of political elections. A trader could purchase a contract through their system (later web-based) which would pay the holder $1 if George H.W. Bush won the 1988 election, and $0 otherwise. This concept showed that markets need not be tied to trading solely physical or financial assets; they can trade futures contracts of any type. These futures contracts, through the market mechanism, evaluated the probability of an event occurring through a price function. If a Bush contract was trading at $0.60, the market evaluated the chance of a Bush re-election at 60%. The principle that the prices can be interpreted as probabilities has been established in theory, with varying approaches on rationality and attitudes toward risk.[2]

Early Iowa Electronic Market results demonstrated that futures contracts for uncertain, non-financial events could be accurate. Even though the profit motive was not as strong (trading was limited to relatively low amounts of money due to U.S. securities regulations), the market was still efficient, and the pricing of contracts, and consequently the probability of a particular event occurring, was accurate.

ASSET STRUCTURE

The market maker needs to decide what information he or she wants from the market so that the asset can be properly designed to elicit that information. Once this decision has been made, the operator can use one of two primary methods of structuring assets. The first is a "winner-take-all" model, where two or more contracts are listed in a market. Of these contracts, one and only one will occur and will thus be cashed-out (typically at either $1 or $100). The market price is then the probability that a particular contract will occur.

Another type of asset is created for use in an index market. That asset cashes out based on the exact state of a particular value at a particular time. For example, an asset can be created that will pay its owner $1 for every percentage point of popular vote won by a presidential candidate. This is the type of futures market found on standard exchanges; examples include the percentage market share achieved or the cost of a barrel of oil to be delivered in 90 days. While other asset structures are certainly possible, a winner-take-all and index contract are the most common in prediction markets.

Elections futures commonly use both types of markets. The first establishes the probability that a candidate will win the election, whilst the second predicts the share of the vote that the candidate will receive.

MARKET STRUCTURE

Within a broad framework of a market, there are a variety of structures that define its operation. The most conventional is the continuous double auction (CDA). In this structure, buyers place "bid" orders at a specific volume and price, while sellers place "ask" orders at a specific volume and price. When the prices match, or the ask price is lower than the bid price, a trade occurs. This continues until all possible orders have been fulfilled and a spread exists again between the bid and ask prices. In some markets, a market maker is used to ensure sufficient liquidity, at a risk of a loss to the market maker. In a CDA, information is continuously updated as traders use news and events to better inform their decisions and make a profit by acting quickly. However, if too few traders exist in a market, the spread between the bid and ask prices can be too large, and thus little or no trading takes place.

An alternate structure is that of a pari-mutuel market. Originally created for sports betting, a pari-mutuel market pools all of the money placed on "bets" for individual contracts, and awards it to the eventual winner, making it appropriate only for winner-take-all markets. The probability of a contract occurring is thus the ratio of the amount bet on an individual contract to the pool of bets in total. Pari-mutuel markets have infinite liquidity, so traders can be guaranteed in this type of market their order will be taken and there is absolutely no risk to the market maker. However, there is no method in which a trader can sell back his or her bet at a profit (assuming the bet was on a contract whose probability of occurring rose). Because of this, there is no advantage for an individual who has already bet to act on new information. Traders should rationally wait until all possible information is known (theoretically just prior to market close) and trade at that time, which means that prices of contracts won't necessarily reflect full information until shortly before market close.

More structures have been developed that try to combine the best aspects of both CDA and pari-mutuel markets. David Pennock of Yahoo Research Labs has created the Dynamic Pari-Mutuel (DPM) market, which aims to be a true hybrid of a CDA and a pari-mutuel market.[3] Robin Hanson has created markets that use Market Scoring Rules (MSR) to adjust prices.[4] While originally developed for use in combinatorial markets to find probabilities of independent variables, an MSR market can also be used for a single variable, such as the probability of an event occurring. Each market structure has benefits that will be attractive to different market operators.

MARKET USES

Prediction markets is one of many terms used to describe using market structures for informative and predictive purposes. The term "prediction markets" reflects their common use: forecasting future events. However, other terms reflect other potential uses: information markets or decision markets

reflect that the prices determined by these markets are used to inform decision-makers and decisions. They can be used in any forum where uncertainty can be properly structured; a film's box office success, an election, or who will win the next series of Big Brother. More importantly for corporations, they can inform business decisions on projected sales of product lines, predict which pharmaceuticals would likely be approved by government regulators, and find efficient solutions to corporate-wide resource allocation problems. To date, prediction markets have been used in resource allocation at BP and in project management for Siemens Germany.[5] As organisations recognise the potential of high-quality information about uncertainty, prediction markets will become more widely adopted.

3. EXPERIMENTAL SETUP

3.1 The problem: predicting race winners

Rowing is a popular amateur sport in England, as well as the greater United Kingdom, as evidenced by the history of the sport in the country. England is home to both Henley Royal Regatta, the oldest running regatta in the world, and the University Boat Race, an annual race between Oxford and Cambridge Universities that was first held in 1829. According to the Amateur Rowing Association, over 20,000 people actively compete each year, who belong to over 550 rowing clubs across England and Wales.[6]

Determining the winners of rowing races is an intriguing problem, as most competitors train in near isolation for over six months from October to April. During the autumn and winter seasons there are only two major long-distance endurance (head) races to gauge a crews' progress relative to the broad spectrum of other club crews. The summer season consists of shorter, Olympic-style sprint races and there are very few regattas that are large, and prestigious, enough to attract top talent from across the country. Most clubs compete in smaller regattas consisting of clubs from their region only. There is no press coverage in the sport on the club-level, nor are there are "experts" or neutral arbitrators that can accurately judge how crews are able to perform against each other before the crews meet in competition.

Particularly at top regattas, where there is a large pool of quality entries, predicting race winners is a difficult task. Injuries, illness, and general crew swapping mean that the squad within a boat frequently changes. Some crews enter multiple events, meaning their performance could be substantially worse for events held later in the day. Many crews have never competed against each other and traders have results from only the head races, a much different style of racing, to use for comparison.

Many of these same issues exist for the national teams competing in World Cup events. Each trains alone for months, emerging only to race in the World Cup regattas, primarily as practice and proving grounds for the World Championships or Olympics each year. Whilst there is more press on the

well-funded national squad crews after a major event, there is little to no information available regarding progress of a crew prior to an event.

Since the problems of crew changes have greater impact and are more frequent at lower levels of the sport, this experiment uses only prediction markets for top-level crews. This was also beneficial from a marketing perspective, as more people would know or recognise the rowers competing at a higher level. More importantly, crew performance at the senior levels of the sport is more consistent. This does not mean it is easier to predict race winners at this level, as the quality of entries is quite high and margin between first and other placings quite small. Stable crews ensure that variations in success or race times are due to the performance of the crew itself, and not from undue influence of other factors.

3.2 Software solution used

This work was undertaken in partnership with Inkling Markets in Chicago, Illinois who ran the website and software for the prediction markets for this research. The primary criteria in choosing a software provider was the user interface; it had to be as easy as possible for novice traders to use, while also providing instant liquidity. These factors were deemed necessary to ensure that new traders would be able to participate immediately after logging into the system. While other software programs provided more sophisticated trading mechanisms, these are best used when users have a strong interest in participation or can receive sufficient training. Inkling's user interface was designed with the most novice participants in mind.

Inkling has three potential algorithms available to market creators: a winner-take-all market, an index market, and a multiple winners market. (The multiple winners market allows for multiple winners in one market, such as the top three out of ten sales categories.) For this experiment the winner-take-all market was selected, which uses a trading algorithm that kept the combined total of all contracts in a market at $100 (100%). As one stock was purchased, its price was subsequently adjusted upward. The software would then adjust all other stocks downward to maintain the total in the market at $100.

A unique feature of Inkling's software is the language and method it uses for new traders to express their judgement. When a particular contract is selected, the trader has three choices. Each choice uses the current price, for example $15 on a $100 scale, and answers the question the market seeks to answer, such as "Will John McCain win the 2008 Republican Presidential Nomination?" The answers are framed: "Chances are higher than 15%," "Chances are exactly 15%," and "Chances are lower than 15%." If a trader selects the "exactly 15%" option, they are informed that they should not trade, and wait until the price changes. The "higher than 15%" and "lower than 15%" choices directs the software to either purchase shares or sell them short, respectively.

The next step of the process is selecting how much to invest in a contract. Instead of purchasing a specific number of contracts at a particular price, as in

a CDA market, each share purchased moves the market price incrementally. Traders are shown four elements: a written expression of their confidence (probability is "Way too low," "Low," or "Way too high," "High," etc.), a number of shares to trade, the amount necessary to make a trade, and an estimated market price after the trade. Depending on the participants' understanding of the market they could trade on any of the three pieces of information.

This step can be very useful as it can work with any mental model of trading that a novice trader holds, and informs them of all the consequences of their actions. (If they trade based on a written description, they are informed how much capital it requires and how high the price will rise or fall as a consequence.) The only potential problem with this implementation is that the software makes an arbitrary link between a written description and its impact on pricing. While there is no real link between the two, it does provide a guide to novice traders that should enable them to accurately express their judgement.

3.3 Trading algorithms

In practice, this trading algorithm is neither a CDA nor a pari-mutuel market. It most closely resembles the Dynamic Pari-mutuel model as theorised by David Pennock, though during this experiment the price function in Inkling's software did not vary with traded volume.[7] Inkling's software has since been updated to use Robin Hanson's Market Scoring Rule, though currently only in one variable.[8] At the time of the experiment, for each share that a trader purchased of a particular contract, the traded price increased by $0.10 no matter the current volume, which means that just one trader can have a significant effect on prices in a given market. This concept and its effects are elaborated further in the section on market manipulation.

The algorithm used allows each individual market to serve as an independent set of trader judgements. While each market's judgements are independent from each other, the set of trader judgements (within each market) are not independent. The algorithm automatically adjusts the prices of the non-traded contracts so that each set's contracts sum to $100.

The following is an example of a market consisting of four contracts using this simple algorithm. Initially the price of each contract is identical, $25:

    A - $25
    B - $25
    C - $25
    D - $25

The first trader decides to purchase shares of contract A, raising the price to $40. The algorithm recalculates the prices for the other three contracts:

A - $40
B - $20
C - $20
D - $20

A second trader then decides to sell shares of contract D short to a value of $5. The prices account for the transaction:

A - $45
B - $25
C - $25
D - $5

A third trader sees the price of contract B, believes it to be undervalued and purchases shares. The algorithm adjusts the prices:

A - $43
B - $31
C - $23
D - $3

This process continues until the market closes. In a thinly traded market, one trader could make a massive purchase that could not be easily countered by others. Conversely, a thickly traded market could see prices swing wildly as significantly biased traders move prices for their favoured contracts in a "ping-pong" effect. These examples are discussed further in a later section.

4. CALIBRATION OF TRADER JUDGEMENT

4.1 Overall calibration

Prediction markets were run for five major regattas: Marlow Amateur Regatta, Henley Women's Regatta, Henley Royal Regatta, Poznan World Cup Regatta, and Lucerne World Cup Regatta. The first three regattas involve competitors from throughout the United Kingdom, while the last two are major international events.

A total of 39 individual markets were created, with one being discarded from calibration results due to incomplete data. A total of 399 competitors (contracts) were entered in the 38 markets. Over the course of the project, 183 traders actively participated in the markets.[9] Figure 1 shows the distribution of popularity of the markets, as measured by how many traders participated.

Because of the high number of contracts in each market, there is a significant concentration of predictions at very low probabilities. With an average of 10.5 contracts per market, a typical contract would begin trading at just above 9.5%. Only three markets were binary options, or between two
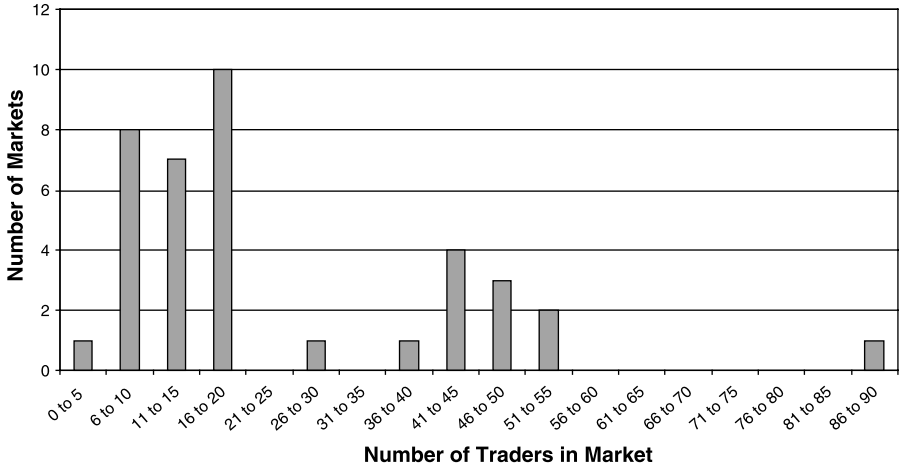
FIGURE 1. Number of markets for each band of traders.

competitors. Many "favourites" in the various markets reached probabilities between 20 to 80%, forcing less popular contracts to probabilities in the 1 to 10% range. Table 1 shows the distribution of contracts in each final traded price band:

A standard calibration of this data, as shown in Figure 2 with identical-width intervals of ten percentage points, shows a relatively poor calibration curve.

This poor calibration is due to the few number of data points in most intervals; over 75% of contracts are in the 0–10% band. This causes the standard regression method to mis-represent the calibration of traders.

The regression is dramatically skewed from perfect calibration by one market: the Stewards' Challenge Cup at Henley Royal Regatta. In this two-boat event, the Great Britain boat was heavily favoured as it had not lost a race

TABLE 1
DISTRIBUTION OF CONTRACTS ACROSS PRICE BANDS

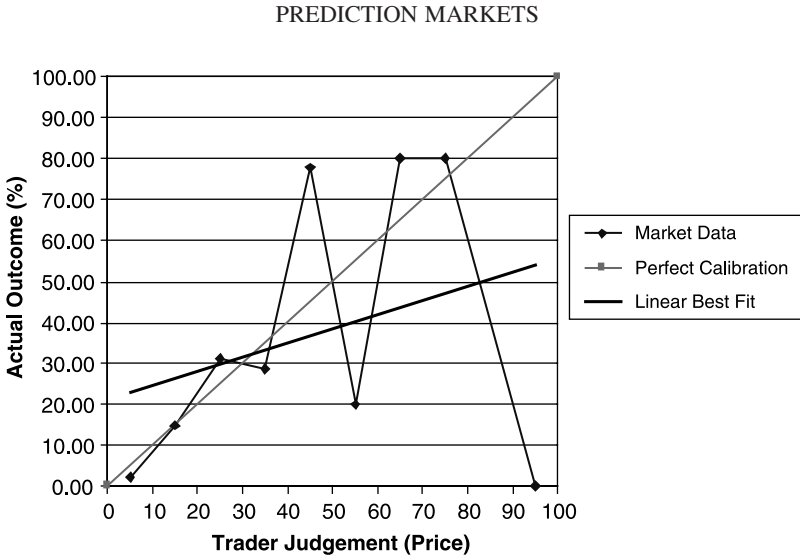| Price Band	# Contracts	# Winners
75 to 100	2	1
65 to 75	8	6
55 to 65	3	1
45 to 55	6	2
35 to 45	8	7
25 to 35	18	5
15 to 25	30	8
5 to 15	81	4
0 to 5	243	4
Total	**399**	**38**

PREDICTION MARKETS

FIGURE 2. Standard calibration of all markets.

in over two years. Consequently, trading in that market pushed its price to just over \$95. However, on the morning of the event illness struck the crew and it pulled out of the race, losing the event. The calibration curve is affected particularly because this race serves as the one and only data point with a trader judgement of over \$80, and thus it radically affects the linear regression. Therefore, the standard calibration model taken without that event is examined, deleting data for the \$95 contract that lost and the \$5 contract that won. This calibration is shown in Figure 3 below:
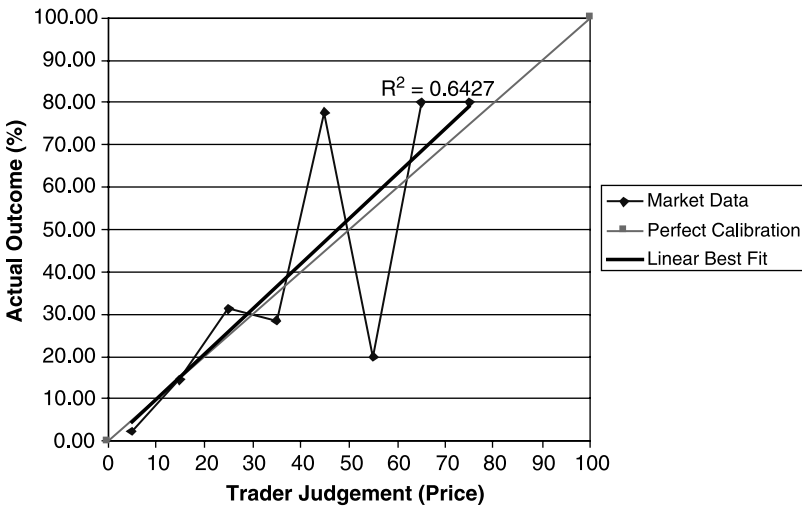

FIGURE 3. Standard calibration of all markets (except Stewards' Challenge Cup).

25

The regression for this data demonstrates nearly perfect calibration. The dramatic swings in the graph are due to having only a small number of data points available to evaluate at higher prices, resulting in a low R^2 value of 0.6427. Without the Stewards' Challenge Cup data points (representing only 0.5% of the total number of contracts) calibration measured in the standard manner is remarkable.

However, a different method can be used to obtain a better representation of the calibration of the market with data points given equal weight. The data was separated into ten intervals with an equal number of contracts in each band. Where a group of identical final prices would be divided into two bands, the number of contracts in each band was adjusted slightly. While this yields irregular intervals, it does ensure enough contracts are in each interval for a proper assessment of calibration. This data set includes the Stewards' Challenge Cup. Using this method, overall calibration of the traders was even more impressive, as shown in Figure 4.

The regression is very nearly perfectly calibrated, particularly at lower probabilities. Having a sufficient number of contracts in each interval causes the R^2 value to increase to 0.9631, an excellent fit. The exact data used is shown in Table 2.

As discussed in the software description, though 399 individual evaluations contributed to this experiment, the data consists of only 38 independent sets of prices. Since the software forces the sum of prices in a market to sum to $100, individual prices cannot be truly independent, though each *set* of prices is independent of each other.
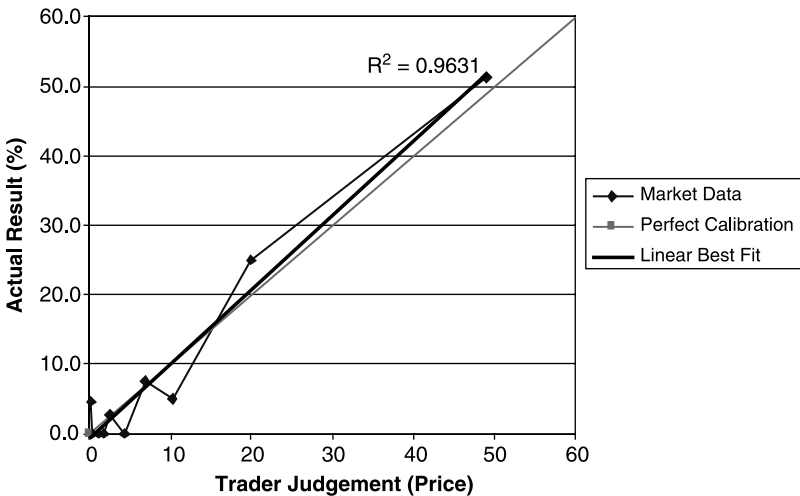

FIGURE 4. Calibration of all markets.

PREDICTION MARKETS

TABLE 2

DISTRIBUTION OF PRICES AS USED IN CALIBRATION CALCULATIONS

Price Band	Average Market Price	# Contracts	# Winners	% Correct
30.58 to 95.22	49.11	39	20	51.3
13.13 to 28.98	20.02	40	10	25.0
8.42 to 12.92	10.2	40	2	5.0
4.78 to 8.25	6.74	40	3	7.5
3.68 to 4.66	4.26	39	0	0.0
1.91 to 3.64	2.48	38	1	2.6
1.75 to 1.89	1.84	40	0	0.0
0.52 to 1.74	1.16	40	0	0.0
0.28 to 0.51	0.38	36	0	0.0
(-)0.45 to 0.27	0.13	44	2	4.5
Total		**399**	**38**	

4.2 Calibration compared to market size

To date, little research has been done into the practicality of small-size prediction markets. Charles Plott and Kay-Yut Chen conducted prediction markets at Hewlett-Packard to predict quarterly sales of various products, monthly sales of products, and profit sharing estimates.[10] They published results based on 12 markets, with between 7 and 26 participants. While they found that their markets were better predictors of future sales than HP official forecasts, there was no assessment of how size affects calibration of the market. Large prediction markets such as NewsFutures, Hollywood Stock Exchange, and Iowa Electronic Markets have proved their efficiency in numerous studies, but few small prediction markets have been assessed.

The number of data points used in these experiments is too small to draw definitive conclusions, but broad trends can be assessed. As shown in Figure 1, there were a number of markets with fairly low participation and a small number of traders, similar to experiments conducted by Plott and Chen. The markets were split into four bands to examine the practical effects of a small number of traders in a market. The number of traders, contracts, and divisions (data points) for each band are shown in Table 3. Figures 5–8 show the calibration of the markets for each range of traders.

TABLE 3

NUMBER OF TRADERS AND CONTRACTS IN EACH DATA Set

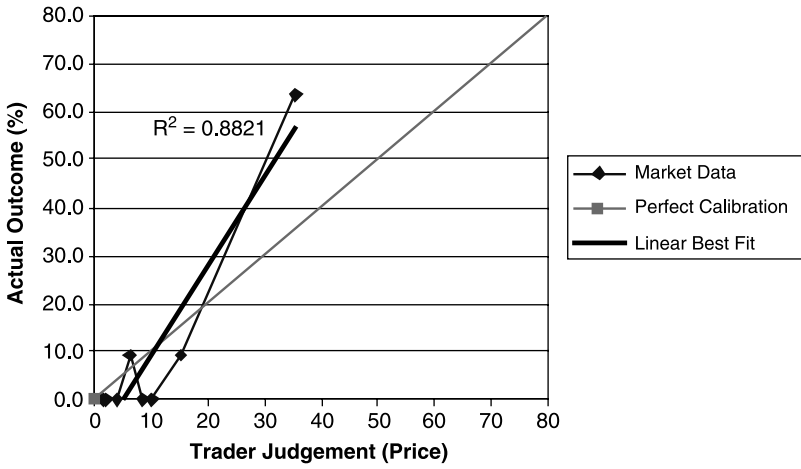| Band	# of Markets	# of Contracts	# of divisions
0–10 Traders	9	89	8
11–15 Traders	7	83	8
16–20 Traders	10	115	10
20+ Traders	12	112	10

FIGURE 5. Calibration of markets with 0–10 traders.

Figures 5–7 show a positive correlation between the quality of the calibration and the number of traders in a market.

In the 0–10 trader band of markets, there is a significant level of under-confidence shown, with contracts at higher prices performing significantly better than what is predicted by the final prices in a market. This improves with additional traders, as seen in Figure 6 with 11–15 traders, though the phenomenon is still present. Figure 7 shows that the under-confidence phenomenon is eliminated and shifts slightly toward over-confidence at higher probabilities when 16–20 traders are participating in the market. Finally, Figure 8 shows that in markets with 21 or more traders, there is an element of
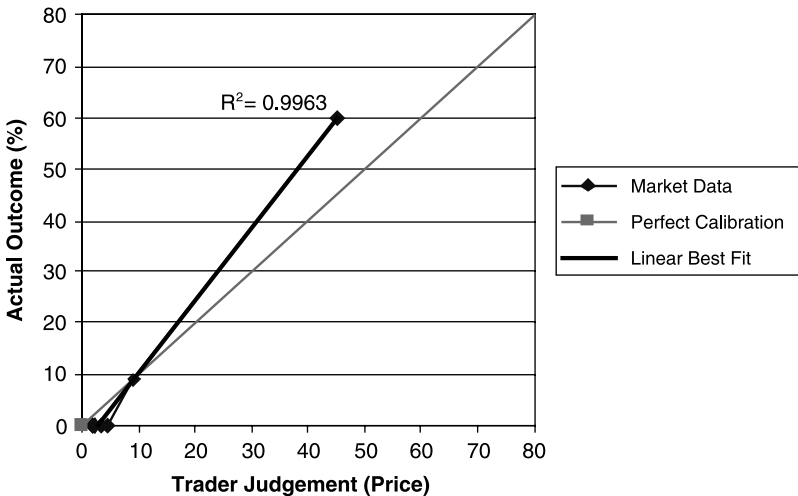

FIGURE 6. Calibration of markets with 11–15 traders.
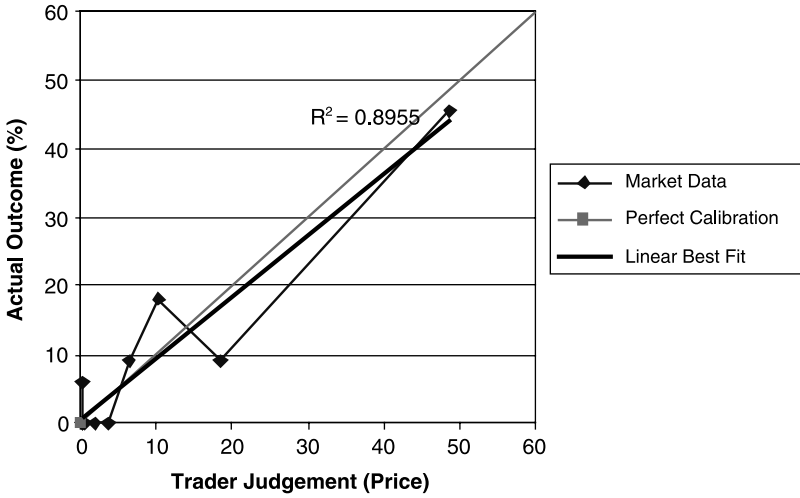
28

PREDICTION MARKETS

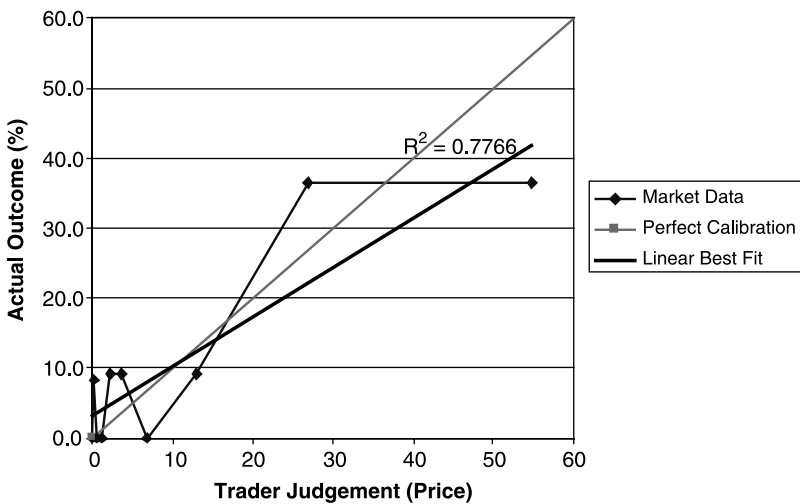
FIGURE 7. Calibration of markets with 16–20 traders.

FIGURE 8. Calibration of markets with 21 or more traders.

over-confidence for contracts at the highest prices, though the fit is poor with an R^2 value of 0.7766.

Economics theory holds that more traders and more activity in a market causes it to become more efficient. In this experiment, markets with more than twenty traders were somewhat less efficient than those with twenty or fewer traders. Again, this data is somewhat skewed because of the inclusion of the Henley Royal Regatta Stewards' Challenge Cup event. Because of the small data set, only 12 independent sets of judgements, this data point is removed because of its inordinate effect on the results. Calibration of the markets
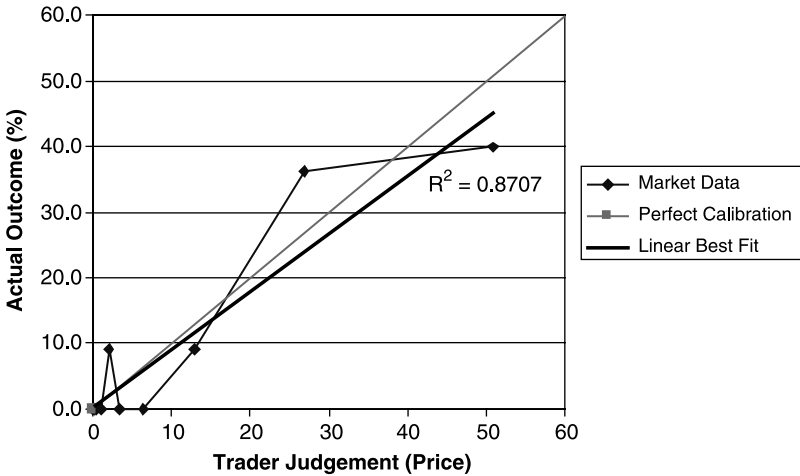
29

FIGURE 9. Calibration of markets with 21 or more traders (exclusive of Stewards' Challenge Cup).

(without the Stewards' Challenge Cup Event) with more than 20 traders is shown in Figure 9 and the difference is clear.

The data without the Stewards' Challenge Cup is more consistent with other markets with higher numbers of traders, such as Figure 7.

The under-confidence phenomenon emerges in markets with a small number of traders, as seen in Figures 5 and 6, due to a combination of the effects of the specific software algorithm used and the general concept of liquidity. An extreme example is the case of the Men's Quadruple Scull (M4x) event at the Poznan World Cup regatta. In this market, only five people chose to trade, and each trader placed one purchase order each. In a standard CDA market, that level of interest would be insufficient to move market prices. Using the algorithm within Inkling's software, there is infinite liquidity so prices did move, though the low level of interest meant prices moved only slightly; not sufficiently enough to truly reflect the consensus judgement of traders. The starting and final prices are shown in Table 4.

The same trend is seen in markets with more traders. For example, the Wyfold Challenge Cup at the Henley Royal Regatta had 12 traders participating. Though there were 32 entries in the event, only 5 contracts saw any trading activity. Of the five contracts traded, four were seeded teams in the event, and the fifth contract consisted of only one trade. The entry with the highest value was Thames Rowing Club 'A' at $42.86, a low figure considering significant wins by the same Thames crew in the previous month at the Metropolitan Regatta and Marlow Regatta, as well as significant interest from the participating traders. Whilst additional traders and interest in this market caused a finer distribution of prices among the favourite crews, there was insufficient interest to create a wider differentiation amongst the less-favourite crews.

TABLE 4

CHANGE IN PRICES IN POZNAN WORLD CUP MEN'S QUADRUPLE SCULLS EVENT

Crew Name	Initial Price	Final Price
Australia	$10.00	$7.75
Belarus	$10.00	$7.76
China	$10.00	$7.75
Czech Republic	$10.00	$15.18
Denmark	$10.00	$7.75
France	$10.00	$7.75
Ireland	$10.00	$7.78
Poland	$10.00	$18.16
Ukraine 1	$10.00	$12.32
Ukraine 2	$10.00	$7.80

In sum, markets with 16 or more traders are shown to be well-calibrated. Markets with 15 or fewer traders are calibrated, but show increasing levels of under-confidence toward the market favourites as fewer traders participate. With a small number of traders, prices simply could not be driven far enough to represent the traders' consensus.

4.3 Calibration of small probabilities

Historically, traders have difficulty properly assessing small probabilities. Studies of transactions in thousands of horse races show that bettors consistently overestimate the probability of unlikely events.[11] As described elsewhere, if small probabilities are biased, they are consistently over-estimates instead of underestimates, particularly in smaller markets.[12]

Justin Wolfers and Eric Zitzewitz write that an open question in prediction markets is: "Are markets well calibrated on small probabilities?"[13] The aggregate data in these markets show that these markets are, in fact, well-calibrated on small probabilities. The prediction markets created for this experiment benefited from a massive number of contracts judged by traders to have a small probability of occurring. There were 243 contracts priced between $0 and $5 (0–5%) at the end of trading. Of these contracts, 4 won their events, for an actual outcome of 1.646%. Whilst the midpoint of the 0–5% interval is 2.5%, the distribution of the contracts within the band is quite uneven; the average contract price in the band is 1.7295%. This is remarkably close to the predicted 1.646%.

5. INCENTIVES IN THE MARKETS

5.1 Participation in rowing markets

The most distinct incentive to attract traders to a prediction market is monetary gain, as can be had through exchanges such as TradeSports, BetFair, Iowa Electronic Markets or similar websites that cost money to trade and

reward traders with real money. Play money exchanges typically use a different kind of reward to encourage participation. This may be t-shirts, company-branded gifts, gift certificates, or similar rewards.

Real-money exchanges are generally believed to be the most accurate, as there can be a significant utility to trading correctly and dis-utility to trading incorrectly. However, recent research has shown that predictions in identical markets on real-money and play-money exchanges are highly correlated with each other, and both are highly accurate.[14] The incentives for trading well in the play-money markets appeared to be a sense of status and some potential prizes.

In the prediction markets operated for this project, there were no incentives provided to traders; no gifts were ever mentioned or promised, no rewards (other than being listed in the top ten) for top traders, or any other prize.[15] In addition to no incentives, the only publicity for these markets was one short post on a rowing news/information/gossip website called The Tideway Slug that is heavily trafficked in the United Kingdom. That post turned viral, with readers informing friends and crew-mates that the website would not have originally reached, with 40% of participants surveyed finding out about the markets from another site, e-mail list, or friend. Though it was only posted to one rowing website, the reach and readership of that website was sufficient to garner a good diversity of traders.

This demonstrates that there may be sufficient incentives to encourage participation without requiring any subsidisation by a market operator. Based on a survey completed at the end of the experiment, there are a few key reasons why this occurred.

COMMUNITY AND UNIQUENESS

Several traders mentioned that they heard about the markets from their friends, and the element of community encouraged participation. Rowing is a relatively small community and many traders in the post-market survey indicated that they liked having something that catered to their interests. When asked about the best aspects of the markets, one participant stated they appreciated "being able to 'bet' on rowing which normally receives no attention," and another remarked "the whole things [sic] was rowing related, it was great."

PERSONAL STAKE

Most traders were eager to participate as they had first-hand knowledge of a crew's performance. In many cases participants were trading on boats in which they were racing.[16] Nearly all traders were and are active participants in the sport and thus are familiar with clubs, boats, and crews from their region.[17] Thus, even if a trader wasn't making predictions on the performance of their

own boat, they were making predictions on boats and crews with which they were familiar.

COMPETITION

Particularly since the traders were competitive athletes, the spirit of competition spilled over into the prediction markets as traders attempted to win in another environment. One participant wrote that the best aspect of the markets were "taking something I knew a little about and trying to apply my knowledge in a competitive way." Other traders saw the markets as a "chance to compete with others" and "fun, and competitive."

5.2 Modelling participation for future events

Lessons from this experiment can potentially inform future construction of incentives in company markets. Whilst the pure passion of a small, competitive sport can inspire participation in a unique forum without any incentives for participation, this is unlikely to be the case in the average business environment.

Unique, interesting events do not require incentives. Yahoo has held periodic brainstorming sessions for new products or services. These ideas, generated by employees, were previously judged by executives to determine the most commercially viable option. Instead of relegating the judging to a small committee of individuals with a similar background and place in the organisation, Yahoo created prediction markets to inform the decision.[18] This is an ideal example of a "unique, interesting event," as the markets could be a fun and popular method for employees to examine each other's ideas and predict which could be successful. Provided there is a low barrier to participation, employees would likely need no incentive to participate.

Certain events where traders feel a very personal stake may also be appropriate for markets without incentives. A potential example of this could be a move or addition to company offices. Markets could indicate which locations employees prefer, while gauging the depth of feeling toward a potential move. The fact that the employee could be personally affected would likely be incentive enough to participate.

6. TRADING BEHAVIOURS

On the individual user level a number of common behaviours emerged from the markets. Over one-third of participants are classified as One-Stake traders, individuals who made one trade in each market in which they participated.[19] Most of these individuals participated in fewer than 7 different markets, though three people traded in as many as eleven markets. They tested the markets, made a small number of purchases, never short sells, and did not make any additional trades based on price movements. Their purchases were essentially votes for the race winner.

Another type of trader is a slightly more complicated version of the One-Stake traders; these are classified as Initial Set traders. The rowing prediction markets yielded 21 traders of this type. These individuals made trades on multiple contracts in a market, and would both purchase shares of some stocks and sell others. However, this type of trader does not sell back (or repurchase in the case of shorts) shares they already hold, so once they express their belief they make no effort to reverse their position. Their purchases and short sells established an initial set of beliefs that did not vary.

The opposite of the One-Stake and Initial Set traders are the Heavy Traders, as they participated in 10 or more markets and made multiple trades per market. There were 12 of these traders; five were in the top ten by net worth at the end of the experiment, with four of the traders in the top five. Seven of these individuals made over 100 individual trades.

The remaining 74 participants can be classified together as Busy Traders, despite the low volume of trades for some individuals. These traders would be willing to sell positions that had risen to ensure a profit, instead of risking the traded boat losing in the event. Occasionally they sold contracts whose value had fallen, presumably for the purpose of limiting potential losses.

In total, 97 of the 183 traders were individuals that applied their beliefs in a single direction; the One-Stake and Initial Set behavioural types. Whether by a lack of interest or lack of knowledge, they purchased a position in each market and didn't revert from it at all. Most of these chose either to purchase or sell only one contract in a market. While just over half of the traders in these markets chose a position and stuck with it, the remainder were responsible for moving prices and causing the variations that existed in the marketplace.

7. MARKET MANIPULATION

Since these markets did not involve real money, there were few barriers to manipulation by traders. In these markets, manipulation was not stopped, but instead let it run its course to determine how the markets were affected. An experimental goal was to assess if price manipulation in a practical prediction market would work, as laboratory-based experiments and some field experiments have shown that market manipulations do not affect price accuracy.[20]

Motivations for market manipulation are varied. In information markets or decision markets, the market may be manipulated to influence the eventual decision being informed. However, the markets in this experiment were designed as prediction markets, where the outcome is based on the quality of forecasts that traders made. Motivations for market manipulation in this case would likely be to make a favoured stock/boat appear much more likely to win for psychological reasons. Since this was a play-money market, another manipulation tactic was observed; participants would register multiple usernames and make trades that would pump money from one username to another.
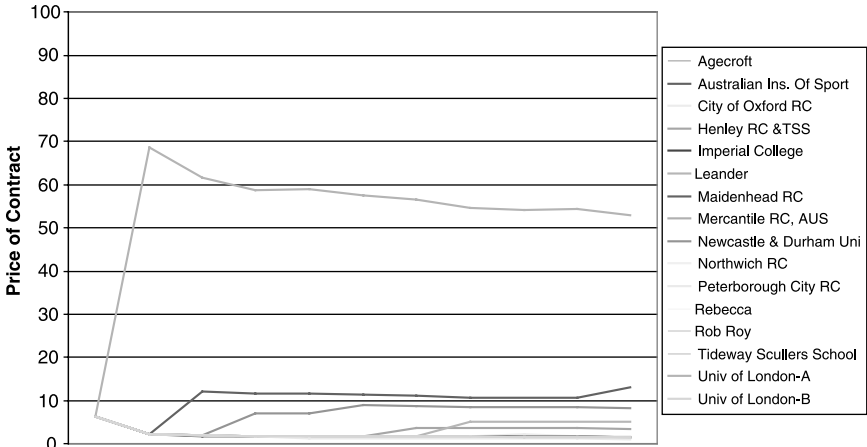
PREDICTION MARKETS

FIGURE 10. Prices over time in Men's Quadruple Sculls. Top line is Univ. of London-A.

 The first tactic appeared to be unsuccessful in markets with many traders, though it was successful in markets with few traders. For example, in the Men's Quadruple Scull event at Henley Royal Regatta (with only 7 traders), the very first trade pushed one entry to a price of $68.75 from it's starting point of $6.25. With only twelve other trades made in the market, there was insufficient liquidity to move that very high starting price. Prices over time are shown in Figure 10. The choice of the particular entry was unusual, as it was not one of the seeded entries at the Regatta.[21] Examining other purchases by the trader showed that the individual bought shares of crews from the University of London, indicating a motive of either enthusiasm or manipulation.

 Any attempts at price manipulation in larger markets were unsuccessful. Nearly every large purchase was then countered, by either the same individual or a competing trader, at some later point by selling back the original purchase. Since the market algorithm used a linear scale to determine the change in price after any purchase, the system is prone to instability in highly contested markets and thus a specific manipulated price is difficult to maintain. Still, the volatility of this algorithm ensures that any manipulation is easily countered. In the market for the Men's Senior 1 Eights event at Marlow Regatta, 89 traders made trades. For the last day of trading, data was obtained showing the price of every contract over time, and is shown in Figure 11.

 Figure 11 shows two distinct manipulations of the market. The first is in favour of Vesta Rowing Club, whose price was driven from near zero to $75 before being traded back to zero. From there it went up to $38 before returning to zero. This series of manipulations and counter-manipulations took place in the span of less than one hours' time. Another major manipulation is shown with traders pushing the price of London RC, the eventual winner, to nearly zero, then purchasing shares to raise the price to $85 before falling to its
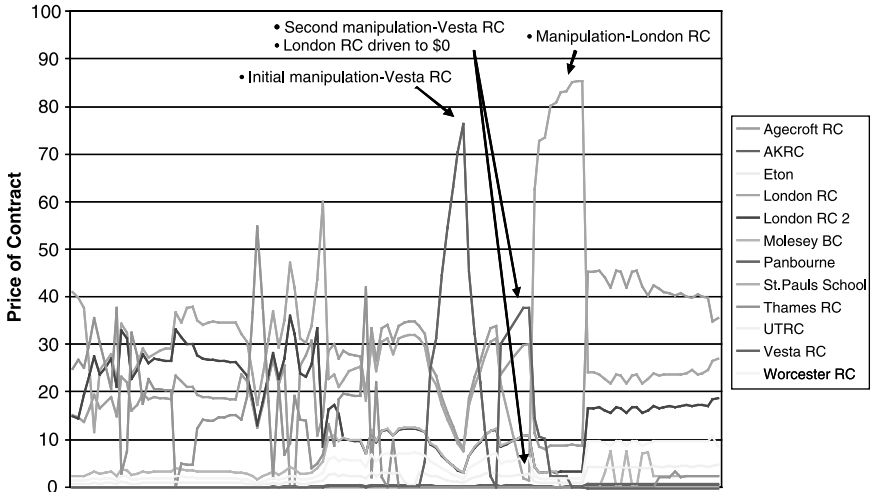
35

FIGURE 11. – Prices over time for Men's Senior 1 Eights at Marlow Regatta, final day of trading.

equilibrium (pre-manipulation) range of approximately $20. This particular series of trades took place in less than twenty minutes.

As discussed previously, unlike the model presented by David Pennock where the price changes for each additional share purchased, the Inkling algorithm was much more basic in it's pricing model.[22] In the markets used for this experiment, each share purchased moved that contract's price by $0.10, no matter the current holdings of the current traders. In smaller markets this model works well as it offers infinite liquidity, but in larger markets will scale poorly as each individual has the ability to single-handedly move prices well outside the market equilibrium. Unfortunately, as Figures 10 and 11 demonstrate, manipulation is much harder to counter in a small market compared to a large market.

The second tactic, registering multiple usernames, was far more common. In a survey completed by 81 of the 183 active traders, four people admitted to registering as multiple users in order to manipulate the markets. Examining registration e-mail addresses, it is estimated that an additional 4–6 traders, some of those that did not answer the survey, used multiple logins to manipulate the markets.

Pumping money from one user account to another was a relatively straightforward process to manipulate the potential capital available to a trader. They would purchase a block of shares in a particular stock then use a different account to purchase additional shares of the same stock, driving the price up even further. The trader would then sell the shares purchased originally at a net profit. Whilst this leaves the second account at a disadvantage, it does improve a trader's standings in his or her original account.

Other traders simply favoured a particular boat in the competition; one trader registered two additional accounts to make more trades in a given market, the Women's Double Scull (W2x) event at the Poznan World Cup. As the chosen stock, New Zealand, was trading at approximately $50 per share, the trader could purchase only approximately 100 shares with each account. By purchasing shares with two different accounts the individual made a significant move in the trading price of that particular stock, though the trader chose not to "lock-in" a profit by selling shares from an original account, nor did the trader use the profits from the eventual wins to make trades in later markets. In that particular case it was a one-time manipulation, and likely a consequence of patriotic fervour.

8. CONCLUSIONS & RECOMMENDATIONS

Prediction markets have been growing in popularity, and have been recognised as a potential decision support tool for businesses and organisations. Through a market structure a company can tap the knowledge and experience of all its employees, regardless of geography, function, or rank within the organisation. Implementing prediction markets within a business for strategic and tactical advantages poses significantly different issues than implementing public prediction markets. The size, scope, and knowledge barriers are very different between a public market and one designed as a strategic decision support tool. The rowing experiments were designed to be reflective of potential real-world prediction markets while examining current topics in prediction market research. A number of principles have been discussed that will prove relevant in future market implementation.

These markets proved to be accurate at aggregating and expressing the opinions of the individual traders involved. The calibration curve of the aggregate data demonstrates prediction markets can accurately assess the probability that a given boat would win an event; in other words, the markets effectively valued the assets involved. While this has been consistently shown to be so in large-scale prediction markets (those with hundreds to thousands of participants) this experiment shows it is also the case in small-scale markets, those with a significantly limited pool of potential traders.

More importantly, the prediction markets analysed in this experiment show that even markets with a very small number of people maintain their ability to accurately assess probabilities. However, the results become somewhat less accurate when there are 15 or fewer traders in a market, and are even less reliable when 10 or fewer traders participate. In the case of this experiment, the price algorithm used in the trading software was a significant factor in producing biased data. When few people participated, the price simply was not able to move far enough to reflect their true judgments. This may be improved by introducing a new pricing algorithm, but is partly a consequence of having a market structure that enables instant liquidity. Future research will be able to measure the difference between the pricing algorithm

at the time of the experiment (a linear function based on number of shares purchased) and the Market Scoring Rules, which were adopted as the pricing algorithm shortly after these markets concluded.

Future prediction markets could ensure higher quality predictions by attracting sufficient numbers of traders. Based on the results presented here, 16 or more traders should be sufficient to obtain quality predictions. Smaller markets may be just as useful, though they may exhibit biases of underconfidence toward market favourites.

Due to the large number of individual contracts in each of these markets, a large number of contracts were assessed with small probabilities. In these experiments the answer to the question, "Are markets well-calibrated on small probabilities?" is conclusively "Yes."[23] However, the rowing prediction markets used n-way claims instead of binary claims. It may simply be easier to assess probabilities in an n-way market.

Insufficient work has been done to examine trader behaviour in prediction markets. This experiment addressed three behavioural topics: incentives for participation, behavioural types, and market manipulation. Understanding these behavioural issues will hopefully shed some light on how even small prediction markets can work so effectively.

Trading in prediction markets is often encouraged by some type of incentive. It is generally understood that for small-scale markets, a certain subsidisation is necessary to ensure sufficient participation. Wolfers and Zitzewitz propose three methods to ensure participation, specifically uninformed traders: offering sports betting, subsidizing the markets, and exploiting career concerns.[24] Whilst this experiment did involve "betting" on sports, no real money was involved, unlike websites like TradeSports or BetFair. Without monetary gain, traders participated purely due to interest and passion in a particular arena. The passion incentive is just as relevant as the sports betting, subsidisation, and exploiting career concerns. While this type of incentive would likely not succeed in markets that predict more mundane topics, such as quarterly sales of product lines, it could be used for special interest events that would then enable traders to become familiar with the concepts and tactics of trading. An interesting market in which traders are passionate about an outcome should not require any incentives at all.

Data analysis in these markets extended to the individual trader level and an attempt was made to model traders' behaviour. The intriguing result of classifying behaviours is that over half of all traders were One-Stake or Initial-Set traders. Traders essentially voted for or against the contracts they thought would win or lose, and didn't make any other trades. A large number of other traders, the Busy Traders at 40% of the trader population, did buy and sell contracts in response to price moves in the various markets. Since the markets were well-calibrated, it demonstrates that even though a large percentage of the trading population did not participate regularly in the form of multiple trades, their behaviour set an initial landscape of pricing in the markets. This landscape was then defined further by the Busy and Heavy Traders.

PREDICTION MARKETS

With the high volume of trades from a competitive population of traders, the markets were sometimes manipulated. The most common method was using multiple user accounts to effectively transfer money from one username to another. The only consequence of this on the markets was a somewhat unfair advantage in capital available to trade. The less common market manipulation was to make large trades to drive prices to extremely high or extremely low levels. This tactic simply did not work in markets with a large number of traders as other individuals rapidly moved prices back to their original levels. The tactic did work in markets with few traders, as there simply were not enough traders to overcome the manipulation.

Overcoming the two manipulation tactics should be relatively straightforward. In a business setting it is easy to ensure that an employee only has one user account, so he or she will be unable to gain additional capital unfairly. Ensuring the market has a sufficient number of traders will eliminate major price fixing through the second form of manipulation. By ensuring that a market has enough traders to be well-calibrated (16 or more) the market will likely have enough participation to counter any manipulation.

Prediction markets, as likely to be used in business environments, will be on a smaller scale than those typically studied due to the significantly smaller pool of potential participants available. The research presented here demonstrates that the principles of prediction markets hold true down to the size that could and would be used in business. Organisations can be assured that using prediction markets, the individual and specialised knowledge of each employee will be included in a company-wide assessment of uncertainty. Whether that uncertainty is forecasting next quarters' sales, the likelihood that project deadlines will be met, or the best location for a new manufacturing plant, a crowd of employees should efficiently assess the uncertainty involved.

NOTES

1. Hollywood Stock Exchange: Pennock, D, Lawrence, S, Giles, C. & Nielsen, F. (2001), Spann, M & Skiera, B. (2003).
   Iowa Electronic Markets: Wolfers, J., & Zitzewitz, E. (2004), Berg, J., & Reitz, T. (2003), Berg, J., Forsythe, R., Nelson, F., & Rietz, T. (2001).
   Tradesports: Servan-Schreiber, E., Wolfers, J., Pennock, D., & Galebach, B. (2004).
2. Manski, C. (2004), Gjerstad, S. (2005), Wolfers, J., & Zitzewitz, E. (2005), Adams, C. (2006).
3. Pennock, D. (2004)
4. Hanson, R. (2003)
5. BP Climate Change Manager Jeff Morgheim (2000), Ortner, G. (1998).
6. E-mail from Tim Page, Amateur Rowing Association Membership Development Manager. (2006 August 8).
7. Pennock, D. (2004)
8. Hanson, R. (2003)
9. A total of 298 individual logins were registered by traders.
10. Chen, K, & Plott, C. (2002).
11. Snowberg, E., & Wolfers, J. (2005).
12. Wolfers, J., & Zitzewitz, E. (2004).
13. Wolfers, J., & Zitzewitz, E. (2006).
14. Servan-Schreiber, E., Wolfers, J., Pennock, D., & Galebach, B. (2004).

15. A prize was promised after all markets were completed to encourage participation in the post-market survey. The prize was rewarded in a random drawing of survey participants.
16. Over 40% of those surveyed.
17. of 81 participants surveyed.
18. King, R. (2006 August 3).
19. 76 of 183 traders.
20. Hanson, R, Oprea, R, & Porter, D. (2005), Hanson, R, & Oprea, R. (2004), Rhode, P. & Strumpf, K. (2005).
21. At the 2006 Henley Royal Regatta, no event with seeded entries was won by an unseeded entry.
22. Pennock, D. (2004).
23. Wolfers, J., & Zitzewitz, E. (2006).
24. ibid.

REFERENCES

C Adams Learning in Prediction Markets (2006) FTC.

J Berg and T Reitz 'Prediction markets as decision support systems' *Information Systems Frontiers* (2003) 5(1) 79–93.

J Berg, R Forsythe, F Nelson and T Rietz 'Results from a dozen years of election futures market research' in Charles Plott and Vernon Smith (eds) *Handbook of Experimental Economic Results* (Amsterdam, Elsevier, forthcoming, 2001).

BP Climate Change Manager Jeff Morgheim 'Statement to the Senate committee on commerce, science, and transportation, full committee hearing on solutions to reduce greenhouse gas emissions' (2000) http://www.gcrio.org/OnLnDoc/pdf/pdf/morgheim000921.pdf. (accessed 1 June 2006).

K Chen and C Plott 'Information aggregation mechanisms: concept, design and implementation for a sales forecasting problem' (2002) California Institute of Technology Social Science Working Paper 1131, March.

S Gjerstad 'Risk aversion, beliefs, and prediction market equilibrium' (2005) Economic Science Laboratory, University of Arizona.

R Hahn and P Tetlock (eds) *Information Markets: A new way of making decisions* (Washington, DC, AEI Press, 2006).

R Hanson 'Combinatorial information market design' *Information Systems Frontiers* (2003) 5(1) 107–119.

R Hanson and R Oprea 'Manipulators increase information market accuracy'. Department of Economics, George Mason University (2004).

R Hanson, R Oprea and D Porter 'Information aggregation and manipulation in an experimental market' (2005) Interdisciplinary Center for Economic Science, George Mason University.

F Hayek 'The use of knowledge in society' *American Economic Review* (1945) 35 519–530.

R King Workers, Place Your Bets, BusinessWeek Online (2006 August 3).

Manski, C 'Interpreting the predictions of prediction markets' (2004) NBER Working Paper No. 10359, March.

Ortner, G 'Forecasting markets- an industrial application' (1998) Mimeo, Technical University of Vienna.

D Pennock 'A dynamic pari-mutuel market for hedging, wagering, and information aggregation' (2004) Yahoo! Research Labs.

D Pennock, S Lawrence, C Giles and F Nielsen 'The power of play: efficiency and forecast accuracy in web market games' (2001) NEC Research Institute Technical Report 2000-168.

P Rhode and K Strumpf 'Manipulating political stock markets: a field experiment and a century of observational data' (2005) UNC Chapel Hill.

E Servan-Schreiber, J Wolfers, D Pennock and B Galebach 'Prediction markets: does money matter?' *Electronic Markets* (2004) 14(3).

E Snowberg and J Wolfers 'Explaining the favorite-longshot bias: is it risk-love, or misperceptions?' (2005) Working Paper.

M Spann and B Skiera 'Internet-based virtual stock markets for business forecasting' *Management Science* (2003) 49(10) 1310–1326.

J Surowiecki *The Wisdom of Crowds* (New York, Anchor Books, 2004).

J Surowiecki Commercial Conversations, Forbes (2005, October 24)

PREDICTION MARKETS

J Wolfers and E Zitzewitz 'Five open questions about prediction markets' in R. W Hahn and P Tetlock (eds) *Information Markets: A New Way of Making Decisions* (Washington, DC, AEI Press, 2006) pp 13–36.

J Wolfers and E Zitzewitz 'Interpreting prediction market prices as probabilities'. Wharton School, University of Pennsylvania (2005).

J Wolfers and E Zitzewitz 'Prediction Markets' *Journal of Economic Perspectives* (2004) 18(2) 107–126.